

ML031272
mat. dossier

(12) UK Patent Application (19) GB (11) 2 363 274 (13) A

(43) Date of A Publication 12.12.2001

(21) Application No 0111024.6

(22) Date of Filing 04.05.2001

(30) Priority Data

(31) 10022520

(32) 10.05.2000

(33) DE

(71) Applicant(s)

Robert Bosch GMBH
(Incorporated in the Federal Republic of Germany)
Wernerstrasse 1, Stuttgart-Feuerbach,
D-70442 Stuttgart, Federal Republic of Germany

(72) Inventor(s)

Mathias Wien

(74) Agent and/or Address for Service

A.A.Thornton & Co
235 High Holborn, LONDON, WC1V 7LE,
United Kingdom

(51) INT CL⁷

H04N 7/26

(52) UK CL (Edition S)

H4F FRG

(56) Documents Cited

US 5907360 A

(58) Field of Search

UK CL (Edition S) **H4F FRG FRP FRX**

INT CL⁷ **H04N 7/26 7/34**

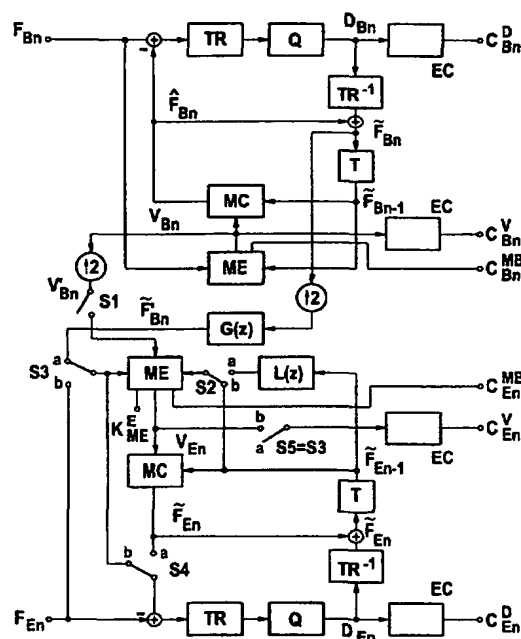
ONLINE: **WPI EPODOC PAJ**

(54) Abstract Title

Spatial scalable moving picture encoding method

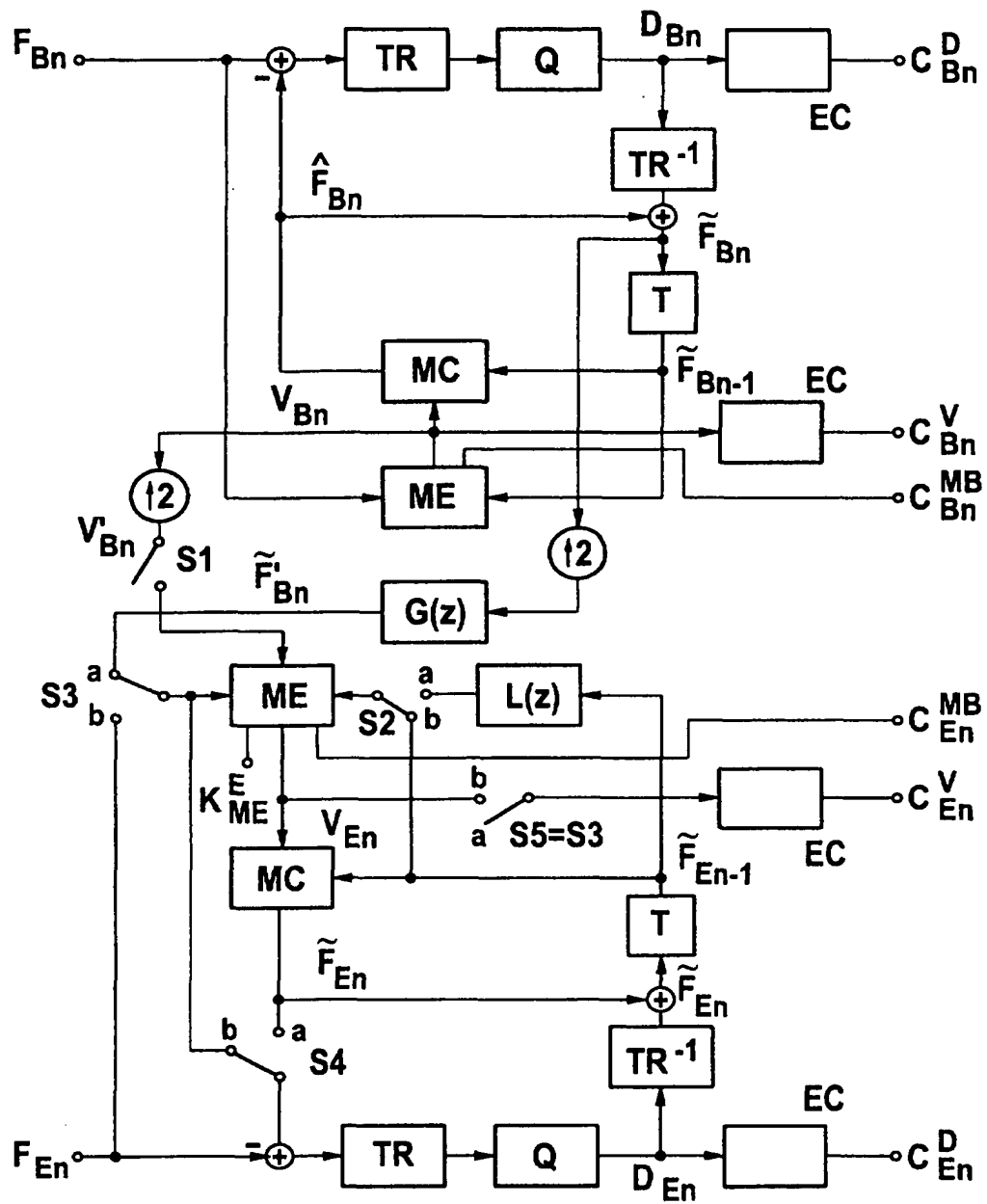
(57) Spatial scalable encoding of a moving picture (e.g. video) is achieved by performing motion estimation (ME) for an image of increased resolution. The high resolution image is based on interpolated versions of a current picture signal and a previously determined or transmitted reference picture signal. Transmission of displacement vectors for the image of increased resolution is therefore unnecessary, allowing the entire data rate to be used for encoding the prediction error.

Fig. 1

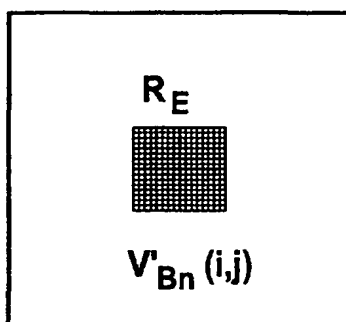


GB 2 363 274 A

Fig. 1



2/2



\tilde{F}'_{Bn}

Fig. 2

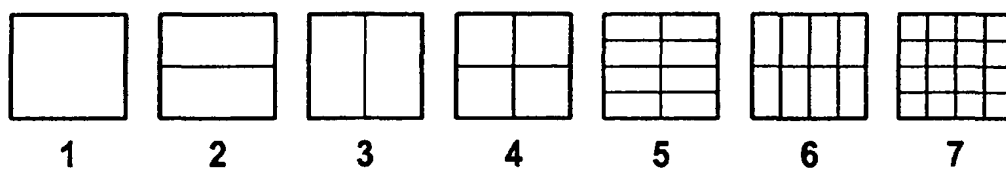
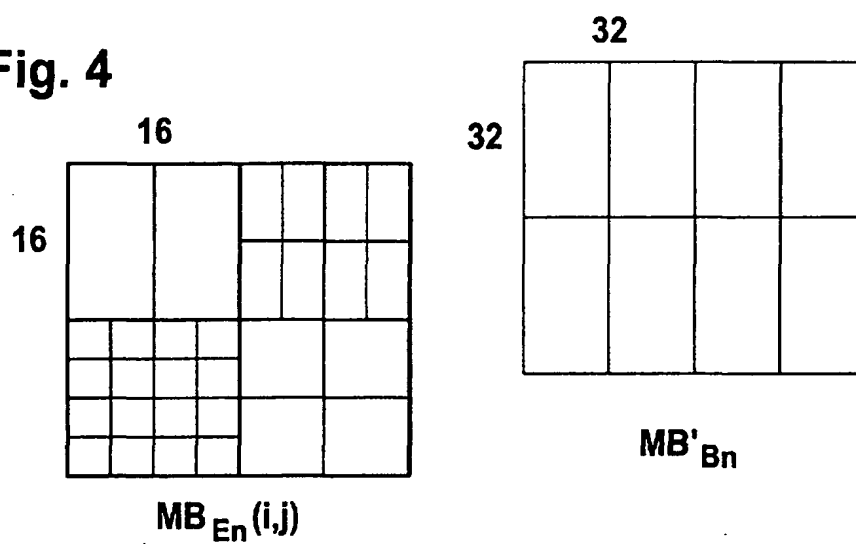


Fig. 3

Fig. 4



5

Spatial scalable moving-picture encoding method

Prior art

10

The invention proceeds from a spatial scalable moving-picture encoding method in at least two stages of different spatial resolution.

- 15 Scalable picture encoding methods make it possible to decode an encoded signal in various resolutions. Normally, the resolution doubles between the scaling stages. To decode a higher resolution, all the lower resolutions are necessary (hierarchical structure). The stages are encoded
- 20 in separate bit streams.

The spatial scalable methods standardized hitherto [1, 2] are based on the hybrid-encoding concept. They have a pyramid structure in which the base layer, i.e. a stage

25 having lower spatial resolution, and the enhancement layer, i.e. a stage having increased spatial resolution, are encoded. To encode the enhancement layer, they use enhanced intraprediction in which no items of information from preceding frames, but from the current base layer,

30 are used, and enhanced interprediction, in which movement vectors and the prediction error are transmitted for the enhancement layer. In this connection, the rate available for the enhancement layer has to be divided up between the movement vectors (displacement vectors) and the prediction

35 error.

In [3], a spatially scalable method is disclosed that manages without the transmission of movement vectors. Here, the prediction is made between two preceding frames and the movement vectors are then extrapolated for the
5 current frame. The term backward motion compensation is used for this method.

In [4] and [5], hierarchical encoding methods are disclosed that are based on discrete wavelet
10 transformation (DWT). In this connection, hierarchical motion estimation is carried out on the hitherto encoded decomposition stages of the DWT of the current and of the reference frames. Since these are known to the transmitter and the receiver, these methods are able to dispense with
15 a transmission of motion vectors.

A single-stage DWT decomposes a frame into row direction and into column direction, in a low-pass (L) component and a high-pass component (H) in each case. This results in
20 four subbands LL, HL, LH and HH, that each have half the row number and column number; the total number of coefficients therefore corresponds to the number of pixels in the frame. In a multistage DWT, said decomposition is applied in each case to the LL band of the current
25 decomposition stage. The LL band is referred to below as the low-pass band and the other bands HL, LH and HH are referred to as high-pass bands.

In the variant proposed in [4], the displacement vectors
30 predicted on the low-pass bands of the coarse decomposition stage of the current and of the reference frames are applied to the high-pass bands of the same decomposition stage. In the case of [5], both low-pass bands of the coarse decomposition stage of the current and
35 of the reference frames are oversampled and interpolated upwards. The predicted displacement vector field is then applied to the low-pass band of the finer decomposition

stage. The single-stage decomposition of this motion-compensated prediction (MCP) is then used as a prediction for the high-pass components of the current frame. In both cases, therefore, predictions are made for the high-pass
5 bands of the coarser stage.

Advantages of the invention

The method of the invention according to Claim 1 and the
10 developments in accordance with the subclaims improves the encoding efficiency of hybrid moving-picture encoded methods having spatial scalability. Said method has the advantage that it is possible to dispense with the transmission of displacement vectors for the stage having
15 increased spatial resolution. The displacement vectors required in the stage of increased spatial resolution EL (enhancement layer) for motion-compensated prediction do not need to be transmitted to the receiver as side information, but are determined at the transmitter
20 (encoder) and at the receiver (decoder) from items of information already known.

Application of the backward motion compensation in the encoding of the enhancement layer avoids a division of the
25 rate between the displacement vectors and the prediction error. The motion estimation is performed on interpolated versions of the current and of the reference frame. Since these are known both to the transmitter and to the receiver, transmission of the predicted displacement
30 vectors as side information is unnecessary, with the result that almost the entire data rate can be used for encoding the prediction error.

The hitherto standardized spatially scalable methods are
35 able to utilize time correspondences only by transmitting displacement vectors. Compared with methods that extrapolate the displacement vectors from preceding

frames, the method according to the invention has the advantage of better correspondence to the motion present in the current frame. Simultaneously, the method can be incorporated well into existing and future standard
5 encoders since no substantial change in the encoder structure has to be made with respect to methods based on DWT.

In contrast to the DWT-based concepts presented at the
10 outset, the enhancement layer is used to predict the displacement vectors in the method according to the invention. It can be optionally low-pass-filtered for the prediction. The method is suitable for block-based application, in particular it can be used in this
15 connection in parallel with the enhanced-intraprediction and enhanced-interprediction methods described above. In methods that permit subdivision of the blocks into subblocks for motion-compensated prediction, the optimum block division can be optionally transmitted by the
20 encoder as side information.

The DWT-based methods are not suitable for application in block-based encoding concepts since block structures in the prediction picture result in the case of DWT in high-
25 pass items of information that are expensive to encode.

Drawings

Exemplary embodiments of the invention are explained in
30 greater detail by reference to the drawings. In the drawings:

Figure 1 shows a block circuit diagram with encoding of the base layer and the facilities for encoding
35 the enhancement layer,

Figure 2 shows the search of the displacement vector for the motion estimation in the enhancement layer,

Figure 3 shows possible divisions of a macroblock,

5

Figure 4 shows the division of four macroblocks of the enhancement layer.

Description of exemplary embodiments

10

Scaling in two stages is described below; the method according to the invention can also be applied analogously to a plurality of scaling stages. The stage having increased spatial resolution is denoted by the enhancement layer (EL) and the stage having lower resolution is called the base layer (BL).

In the method according to the invention, the current BL frame already transmitted is set to the size and resolution of the EL frames by increasing the sampling rate and interpolation filtering. As a reference, use is made of the preceding picture frame of the EL, which is already available to the encoder and decoder. Optionally, the reference frame may be low-pass-filtered so that it does not contain any higher frequency components than the corresponding upward-interpolated BL frame. A motion estimation is performed between the upward-interpolated BL frame and the reference frame. Since the frames used are known to the transmitter (encoder) and to the receiver (decoder), the motion estimation can be performed both at the encoder and at the decoder so that transmission of the predicted displacement vectors is unnecessary. The displacement vectors are used for motion-compensated prediction MCP of the current EL frame to be encoded. The preceding EL frame, which may likewise optionally be low-pass-filtered beforehand, is again used as reference in the motion-compensated prediction MCP. In such encoding

methods, which permit the subdivision of a block into subblocks of various sizes in the motion-compensated prediction MCP, the optimum division of the EL blocks into subblocks can optionally be determined at the encoder and
 5 transmitted as side information to the receiver.

The method of the invention can optionally be used either for all the blocks of the EL frame to be encoded or can be used as an alternative to the MCP modes already provided
 10 in the encoding method.

The method according to the invention is explained below on the basis of the exemplary embodiment of the luminance component of a picture sequence. The encoding is to take
 15 place in a block-oriented manner on the basis of so-called macroblocks (MB) comprising 16×16 pixels.

The method according to the invention shall be denoted by EBP (enhanced backward prediction). The interprediction
 20 hitherto used is denoted by EFP (enhanced forward prediction) and intraprediction as EIP. The enhancement layer shall be greater by a factor of 2 than the base layer in the horizontal and vertical directions. This size ratio is normally used; other size ratios can likewise be
 25 implemented.

The n th frame of a picture sequence is denoted by F_n . The symbol V_n is used for the motion vector field, while the quantized prediction error is D_n . \hat{F}_n denotes a prediction
 30 of F_n , whereas the reconstruction is represented by \tilde{F}_n . The indices B and E denote, respectively, the base layer and the enhancement layer of the corresponding frame. A macroblock is denoted by MB and a subblock of the macroblock by B. The upward-interpolated version of the
 35 frame is denoted by F'_n and the scaled version of the

motion vector field by V'_n . \hat{F}_n is a low-pass-filtered version of F_n .

In the description, the reference frame is characterized
 5 by F_{n-1} , which refers to the preceding frame in time. A frame at another time interval or a selection of preceding frames can also be used as a reference for the prediction.

C_n^D , C_n^v and C_n^{MB} denote the encoded prediction errors, the
 10 motion vectors and the information for dividing up a macroblock MB. The costs K_{ME} that arise in the motion estimation are made up of the sum of the absolute differences SAD between the current and the shifted reference block and, optionally, the costs for the
 15 encoding, for example vectors, block division.

Figure 1 shows a simplified block circuit diagram having base-layer and enhancement-layer encoding. The encoding of the base layer corresponds to the known hybrid encoding
 20 concept, such as is used in principle in the established standards; it is explained briefly here in order to introduce the notations used. Forward prediction \hat{F}_{Bn} is made for the current base-layer frame F_{Bn} by motion estimation ME and motion compensation MC from the reference frame \tilde{F}_{Bn-1} .
 25 The resultant motion vector field V_{Bn} is entropy-encoded EC and transmitted to the receiver. The search area in compensation with 16×16 blocks may be set, for example, at 16 pixels in each direction.

30 The prediction error between F_{Bn} and \hat{F}_{Bn} is transformed (TR, for example by means of the discrete cosine transformation DCT) and quantized. This quantized difference signal D_{Bn} is, on the one hand, encoded and transmitted to the receiver and, on the other hand,

inverse-transformed by means of TR^{-1} and added to the prediction \tilde{F}_{Bn} , resulting in a reconstructed frame \tilde{F}_{Bn} at the receiver. The latter is temporarily stored in a buffer T in order to serve as reference \tilde{F}_{Bn-1} for the next frame. Q denotes the quantizer.

The method is applied macroblock-wise. If various modes, for example intra or inter, or divisions, are provided for the macroblocks, these have to be transmitted additionally as side information C_n^{MB} . The possible entropy encoding for C_n^{MB} , like the choice between intra-encoding and inter-encoding is not shown in the block diagram for reasons of clarity.

15 Basic method

Initially, the switches in Figure 1 are set as follows: S1 = open, S2 = b, S3 = a, S4 = a. Since switches S5 and S3 are coupled, no displacement vectors are transmitted in this case. Let the switch positions be fixed. V_{En} is estimated between the base-layer frame \tilde{F}_{Bn} , upward-interpolated by oversampling and filtering with the interpolation filter $G(z)$, and the enhancement-layer reference frame \tilde{F}_{En-1} .

The motion estimation ME estimates the motion for the current block. This can be performed in the form of a dense displacement vector field or in a block-based manner. A displacement vector field is referred to as dense if a separate vector exists for every pixel of the compensated area. In the case of block-based methods, a common vector is allocated to a block, for example 8×8 pixels. No vectors and, in the block-based case, no items of information about the block division are transmitted.

Filtering of the reference frame

For this purpose, the switch S2 is set to position a. Let
 5 the switch positions be fixed, as in the basic method. V_{En}
 is now estimated between the upward-interpolated base-
 layer frame \tilde{F}'_{Bn} and the enhancement-layer frame \tilde{F}^*_{En-1} low-
 pass-filtered by $L(z)$. The purpose of the filtering is to
 match the frequency response of the reference frame to
 10 that of the upward-interpolated base-layer frame.

Simplified vector search

For this purpose, switch S1 is closed. As a result, V'_{Bn} is
 15 applied to the motion estimation block ME of the enhanced
 layer EL and serves to initialise the vector estimation.
 The prediction vector field V'_{Bn} is produced by scaling V_{Bn}
 by the factor 2 and is consequently matched to the size of
 the enhancement layer. The search is performed in a
 20 reduced search area around the scaled base-layer vector,
 for example two pixels, in order to minimize the search
 expenditure. This is shown in Figure 2. Around the scaled
 motion vector $V'_{Bn}(i, j)$, the search is performed on the
 interpolated frame \tilde{F}'_{Bn} with a reduced search area R_g .

25

Transmission of the block division

In order to minimize the search expenditure for the motion
 estimation at the decoder end, in the block-based method,
 30 the subdivisions of the macroblocks MB can permit C_{En}^{MB} to
 be transmitted as side information. The search for the
 vectors has then to be performed only for the block
 division already transmitted.

35

Choice of the prediction mode

In this operating mode, the method according to the invention is used in parallel with the known prediction
 5 modes. For this purpose, the encoding costs for EIP (S1 = open, S2 = b, S3 = a, S4 = b), EFP (S1 = open, S2 = b, S3 = b, S4 = a) and EBP (switch positions as described above) are compared and the most favourable method is chosen for every macroblock MB.

10

Use of different block sizes.

The possible subdivisions of the macroblock are in line with the subdivisions proposed in the test model TML-3 for
 15 video encoding standard H.26L [6]. The macroblock can be decomposed into subblocks in the manner shown in Figure 3, thereby resulting in subblocks of sizes 16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 and 4×4 pixels. In the enhancement layer, four macroblocks correspond to an
 20 upward-interpolated base-layer macroblock. The macroblock division used in the base layer is enlarged in \tilde{F}'_{Bn} by a factor of 2 as a result of the interpolation. The size of the subblocks in the enhancement-layer macroblocks must not exceed this base-layer division since block artefacts
 25 may otherwise occur within the enhancement-layer blocks.

In Figure 4, which shows diagrammatically the division of four macroblocks $MB_{En}(i, j)$, where $i, j = \{0, 1\}$ of the enhancement layer as a function of the division of the
 30 corresponding interpolated base-layer macroblock MB'_{Bn} , four possible divisions are shown for enhancement-layer macroblocks if the division 6 in Figure 3 has been chosen in the corresponding base-layer macroblock.

35 The division should be chosen for the enhancement-layer macroblocks in such a way that the prediction error to be

encoded is as small as possible. For this purpose, the motion vectors determined are applied to the unfiltered enhancement-layer signal \tilde{F}_{En-1} , and the most favourable block division is transmitted to the receiver as forward
 5 information.

The method according to the invention is suitable for application in spatially scalable picture sequence encoding using H.26L.
 10

For macroblocks that have been encoded with EBP, the latter has to be signalled in the macroblock header, but no motion vectors are encoded in addition.

15 Literature

- [1] ISO/IEC JTC1 IS 14496-2 (MPEG-4). "Information technology - generic coding of audio-visual objects (final draft of international standard)", Oct. 1998.
 20
- [2] Telecom. Standardization Sector of ITU, "Video coding for low bitrate communication (H.263 Version 2)", Sept. 1997.
- [3] T. Naveen and J. W. Woods, "Motion compensated multiresolution transmission of high definition video", IEEE Trans. on Circuits and Systems for Video Technology, vol. 4, pp 29-41, Feb. 1994.
 25
- [4] A. Nosratinia and M. T. Orchard; "Multiresolution backward video coding", in Proc. IEEE Int. Conf. Image Processing ICIP '95, vol. 2, pp. 563-566, Oct. 1995.
 30
- [5] X. Yang and K. Ramchandran, "Hierarchical backward
 35 motion compensation for wavelet video coding optimized

interpolation filters", in Proc. IEEE Int. Conf. Image Processing ICIP '97, vol. 1, pp. 85-88, Oct. 1997.

- [6] Telecom. Standardization Sector of ITU, "H.26L test
5 model long term 3", in Study Group 16, Question 15,
Meeting J. (Osaka, Japan), ITU, Mar. 2000.

5

Claims

1. Spatial scalable moving-picture encoding method in at least two stages (EL, BL) of different spatial resolution
10 employing the following procedure:
 - the motion estimation (ME) is performed for a stage (EL) of increased spatial resolution on the basis of interpolated versions of a current picture signal and of a reference picture signal, wherein a picture signal
15 determined previously in time or transmitted is used as reference picture signal.
2. Method according to Claim 1, characterized in that the displacement vectors for the stage of increased
20 spatial resolution are determined at the encoder end and decoder end from items of information already known and, consequently, do not have to be transmitted as side information to the decoder.
- 25 3. Method according to Claim 2, characterized in that the encoding expenditure saved by nontransmission of the side information is essentially used to encode the prediction error.
- 30 4. Method according to one of Claims 1 to 3, characterized in that a current picture signal, already transmitted, of the stage (BL) of low spatial resolution is set to the size and resolution of the stage (EL) of increased resolution by increasing the sampling rate and
35 interpolation filtering and is compared with the reference picture signal of the stage (EL) of increased resolution for the purpose of motion estimation.

5. Method according to one of Claims 1 to 4, characterized in that the reference picture signal is low-pass-filtered.
- 5
6. Method according to one of Claims 1 to 5, characterized in that the displacement vectors are used for motion-compensated prediction (MCP) of the current picture signal, to be encoded, of increased resolution.
- 10
7. Method according to Claim 6, characterized in that a picture signal determined previously in time or transmitted is used as reference for motion-compensated prediction.
- 15
8. Method according to one of Claims 1 to 7, characterized in that the motion estimation (ME) is undertaken in a block-based manner.
- 20
9. Method according to one of Claims 1 to 8, characterized in that a parallel application is undertaken with enhanced-intraprediction and/or enhanced-interprediction methods.
- 25
10. Method according to Claim 8 or 9, characterized in that, for a subdivision of blocks into subblocks, the optimum block division is transmitted as side information to the receiver.
11. Method substantially as hereinbefore described with reference to the accompanying drawings.



INVESTOR IN PEOPLE

Application No: GB 0111024.6
Claims searched: 1-11

Examiner: Iwan Thomas
Date of search: 4 October 2001

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.S): H4F FRP, FRG, FRX

Int Cl (Ed.7): H04N 7/26, 7/34

Other: Online: WPI, EPODOC, PAJ

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
X	US 5907360A (Kessler) See abstract and col. 3 lines 10-33	1

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.